DATS 6312 : NATURAL LANGUAGE PROCESSING

INDIVIDUAL PROJECT REPORT

ON

CUSTOMER COMPLAINT ANALYSIS AND PREDICTION SYSTEM

Leveraging NLP To Enhance Customer Experience

BY:

Swathi Murali Srinivasan (G29020307)

Overview of this project

1. **Sentiment Analysis of Complaints:**

   o Analyzes the sentiment of customer complaints.

   o Helps companies prioritize and escalate complaints faster based on sentiment.

2. **Complaint Summarization:**

   o Summarizes complaints concisely, especially when they are lengthy or written in poor English.

   o Assists customer service teams in understanding complaints more effectively.

3. **Resolution Prediction:**

   o Predicts the likely resolution for complaints based on historical data.

   o Provides customers with insights into possible outcomes, improving transparency and customer satisfaction.

4. **Complaint Classification:**

   o Classifies complaints into specific categories.

   o Enables companies to route issues to the appropriate teams, ensuring quicker resolution.

Data Used:

Consumer Complaint Database by CFPB: https://www.consumerfinance.gov/data-research/consumer-complaints/

Rows: 2199541

Size of Data: 4.05GB

Dates: 01/01/2018 to 11/11/2024

Below are the topics which I have worked on

1. Exploratory Data Analysis
2. Sentiment Analysis
3. Streamlit Application for Sentiment Analysis

1. Exploratory Data Analysis:

Data Overview

The dataset from the Consumer Financial Protection Bureau (CFPB) comprised 2,199,541 rows and was 4.05 GB in size, containing complaint data from **01/01/2018 to 11/11/2024**. Key columns analyzed included:

- **Product and Sub-product**: Identifying the type of financial products involved in complaints.
- **Issue and Sub-issue**: Categorizing customer concerns.
- **Timely Response**: Whether the company responded to the complaint promptly.
- **Company**: Names of companies receiving complaints.

Insights from Data Analysis

1. **Top Products by Complaint Volume**:
   - The products with the highest complaints included "Debt collection," "Mortgage," and "Credit reporting."
   - Sub-products provided further granularity, revealing nuances such as "Conventional fixed mortgages" in mortgage-related complaints.
2. **Timely Responses**:
   - Approximately 90% of the complaints received a timely response, reflecting the regulatory pressure on companies to resolve complaints quickly.
3. **Complaint Distribution by Year**:
   - A steady increase in complaints was observed year over year, potentially linked to increasing consumer awareness of the CFPB platform.
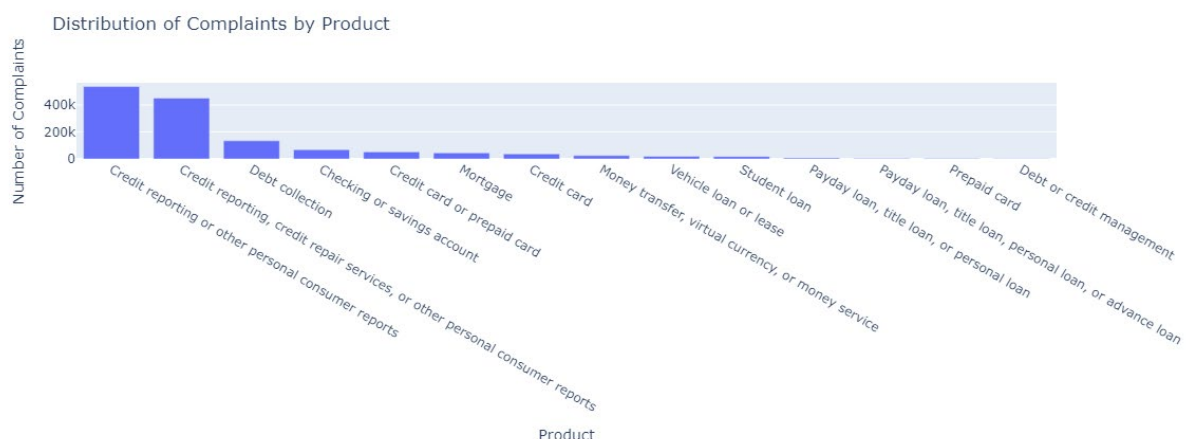4. **Word Count Analysis**:
   - Processed narratives ranged from brief complaints (~10 words) to detailed submissions (~500 words).
   - Average word count was 148, with longer complaints often corresponding to complex issues such as mortgage disputes or fraud cases.
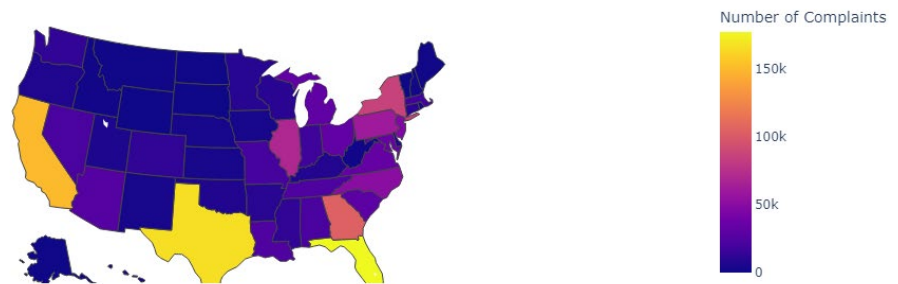
Visualizations

Key plots generated:
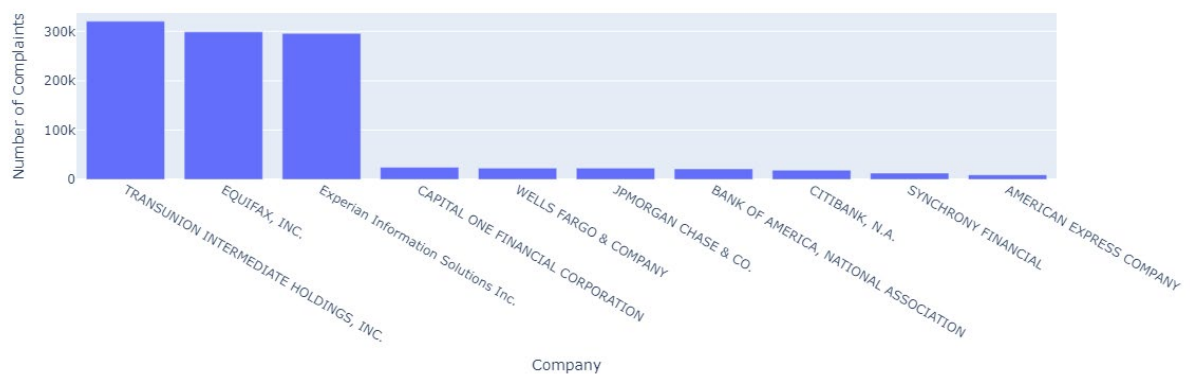
1. Product-wise Complaint Volume

2. Complaint Count – State-wise

State-wise Complaint Count

3. Top 10 Companies with most complaints:

Top 10 Companies with the Most Complaints

2. Sentiment Analysis:

Task Overview:

My goal was aimed at developing an accurate and efficient sentiment analysis model to classify customer complaints into three sentiment categories: Positive (2), Neutral (1), and Negative (0). The work involved meticulous preprocessing, model selection, training, and evaluation to ensure reliable results that could enhance customer service decision-making processes.

Initial Model Comparison:

Data Preparation and Sentiment Labeling:

The dataset contained customer complaint narratives that were cleaned and preprocessed in earlier steps. The "Processed Narrative" column served as the input text for model training and evaluation. Sentiment labeling was conducted using the VADER Sentiment Intensity Analyzer, which assigns a sentiment score to each narrative based on polarity:

Compound Score ≥ 0.05: Positive sentiment

Compound Score ≤ -0.05: Negative sentiment

Otherwise: Neutral sentiment

This labeled dataset was split into 60% training, 20% validation, and 20% test sets to ensure robust evaluation. A stratified sampling approach was used to maintain class distribution across splits.

Models Compared:

To identify the optimal model for sentiment classification, several pre-trained transformer architectures were fine-tuned and evaluated:

1. **BERT (base-uncased):** A general-purpose transformer widely recognized for its robust performance across various NLP tasks.
2. **RoBERTa (base):** An optimized version of BERT that enhances training by improving masking strategies and using larger datasets.
3. **DistilBERT (base-uncased):** A compact, faster, and efficient variant of BERT, suitable for resource-constrained environments.
4. **T5 (base):** A sequence-to-sequence transformer that can adapt to a variety of tasks, including sentiment classification.

Evaluation Metrics

The models were evaluated on the following metrics to provide a holistic performance overview:

- **Accuracy**: The ratio of correctly classified instances to the total number of instances.

- **Precision**: The proportion of true positive predictions to all positive predictions, reflecting model specificity.

- **Recall**: The proportion of true positive predictions to all actual positives, reflecting sensitivity.

- **F1-Score**: The harmonic mean of precision and recall, balancing these metrics for overall effectiveness.

Results:

Model	Accuracy	Precision	Recall	F1-Score
BERT (base-uncased)	0.9028	0.84	0.82	0.83
RoBERTa (base)	0.8920	0.88	0.80	0.83
DistilBERT (base-uncased)	0.8942	0.84	0.81	0.82
T5 (base)	0.8418	0.81	0.68	0.72

Analysis of Model Comparison Results:

1. **BERT (base-uncased)** emerged as the best-performing model, achieving the highest **accuracy** (0.9028) and **F1-score** (0.83). Its balanced **precision** (0.84) and **recall** (0.82) indicated consistent performance across all sentiment classes.

2. **RoBERTa (base)** demonstrated the highest precision (0.88), making it highly effective in reducing false positives. However, its recall (0.80) was slightly lower, suggesting a tendency to miss some true positives.

3. **DistilBERT (base-uncased)** delivered competitive results, with only marginal differences from BERT. Its lightweight design makes it a strong contender for applications requiring faster inference times.

4. **T5 (base)** underperformed relative to the other models, particularly in recall (0.68), indicating challenges in generalizing to the dataset.

Conclusion

Based on the evaluation, **BERT (base-uncased)** was selected for fine-tuning and deployment due to its superior overall performance. Its ability to balance precision and recall while achieving high accuracy makes it well-suited for the nuanced task of sentiment analysis in customer complaints.

Model Implementation:

Model Selection:

After comparing multiple models, **BERT (base-uncased)** was selected as the best-performing model based on its high accuracy, precision, recall, and F1-score across all sentiment categories.

Data Preparation for Training

1. **Stratified K-Fold Cross-Validation**:

   o   Implemented 5-fold cross-validation using StratifiedKFold from scikit-learn.

   o   This ensured balanced splits based on sentiment class distributions.

2. **Custom PyTorch Dataset**:

   o   Texts were tokenized using BertTokenizer.

   o   A PyTorch dataset class was created for dynamic data batching.

Model Training Process:

The model was fine-tuned using the following configuration:

- **Model**: BERT (bert-base-uncased)

- **Batch Size**: 32

- **Learning Rate**: 2e-5

- **Epochs**: 5

- **Optimizer**: AdamW with linear warm-up scheduler

Training Techniques:

- **WeightedRandomSampler**: To address class imbalance.

- **Gradient Clipping**: Prevented exploding gradients.

- **Learning Rate Scheduler**: Improved training stability.

For each epoch, training loss and validation loss were logged.

Evaluation Metrics

- **Accuracy**: Overall correct classifications.

- **Precision**: TP / (TP + FP)

- **Recall**: TP / (TP + FN)

- **F1-Score**: Harmonic mean of precision and recall.

Results and Performance Analysis

Class-wise Metrics

	Precision	Recall	F1-Score	Support
Negative	0. 97	0.97	0. 97	31,258
Neutral	0.94	0.92	0.93	3,342
Positive	0.97	0.97	0.97	33,048

Overall Performance:

- **Accuracy**: 0.97

- **Macro Average**: 0.96

- **Weighted Average**: 0.97

Model Deployment:

Sentiment Prediction Module

A prediction module was developed for batch inference using pre-trained BERT models.

1. **Model Loading**:

   o The saved model and tokenizer were loaded using Hugging Face's from_pretrained() function during the deployment phase.

2. **Prediction Process**:

   o Tokenized input texts with padding and truncation.

   o Inferred sentiment class labels using a softmax classifier.

   o Results were mapped back to sentiment labels (Negative, Neutral, Positive).

Streamlit App Integration

- A Streamlit application was developed to demonstrate the model's inference capabilities.

- **Application Workflow**:

   o Users can enter complaint texts into a text box.

   o Upon submission, the text is tokenized and passed through the trained model.

   o The model's prediction is displayed with corresponding sentiment indicators.

Inference Batch Processing

- Batch size: 32

- GPU acceleration was enabled with CUDA, clearing the cache after each batch inference.

Key Insights and Contributions

- **Data Cleaning and Preprocessing**:

   o Designed and verified the entire preprocessing pipeline.

- o Implemented consistency checks, ensuring a clean, well-processed dataset.
- **Model Training and Evaluation**:
  - o Designed the model training loop with 5-fold cross-validation.
  - o Integrated advanced training techniques like weighted sampling and learning rate scheduling.
- **Model Deployment**:
  - o Saved trained models for deployment.
  - o Developed a robust prediction pipeline for new incoming data.
  - o Integrated the model into a Streamlit-based interactive demo application.

Test Cases: Application Results

Since our model was not trained on the data before 2018, we tested it with CFPB complaints predating that year.

Test Case 1:

2765408

Date CFPB received the complaint
12/27/2017

Consumer's state
NJ

Consumer's zip
085XX

Submitted via
Web

Did consumer dispute the response?
N/A

Product
Credit reporting, credit repair services, or other personal consumer reports
Sub-product: Other personal consumer report

Issue
Problem with a company's investigation into an existing issue
Sub-issue: Investigation took more than 30 days

Consumer consent to publish narrative
✓ Consent provided

Consumer complaint narrative
Hi, Company Name : Paypal Dispute transaction date : XX/XX/XXXX in the amount of {$500.00} (Fraud transaction) Dispute filed : XX/XX/XXXXDispute ID : XXXX Dispute description : Unauthorised transaction Initial decision from Paypal received on XX/XX/XXXX was to deny my claim After that I appealed to Paypal fraud specialist on XX/XX/XXXX, and I got this message : " Hello XXXX XXXX, We recently received your appeal for an unauthorized case that was denied. We 'll get back to you with our findings within 9 days. '' Its been more than 75 days and I have n't received any update or Credit of {$500.00} During this time, I have contacted them twice and both times they have told me that the case is still under review and will be expedited. Nothing is happening. Paypal should immediately credit {$500.00} to my account. Thanks XXXX XXXX

Company information

Date complaint sent to company
12/27/2017

Company name
Paypal Holdings, Inc

Timely response?
◉ Yes

Company response to consumer
Closed with monetary relief

Company public response
N/A

The above data is from CFPB on date 12/27/2017

App Response:

Customer Complaint Sentiment and Summary Analyzer

Enter your complaint text below to analyze its sentiment and generate a summary.

Enter the Consumer Complaint Narrative:

appealed to Paypal fraud specialist on XX/XX/XXXX, and I got this message : " Hello XXXX XXXX, We recently received your appeal for an unauthorized case that was denied. We 'll get back to you with our findings within 9 days. " Its been more than 75 days and I have n't received any update or Credit of {$500.00} During this time, I have contacted them twice and both times they have told me that the case is still under review and will be expedited. Nothing is happening. Paypal should immediately credit {$500.00} to my account. Thanks XXXX XXXX

`Analyze`

Sentiment Analysis

Sentiment: Negative

Summary

Generated Summary: Initial decision from Paypal received on XX/XX/XXXX was to deny my claim. After that I appealed to Paypal fraud specialist on XX/XX/XXXX, and I got this message : " Hello XXXX XXXX, We recently received your appeal for an unauthorized case that was denied. We 'll get back to you with our findings within 9 days " Its been more than 75 days and I have n't received any update or Credit of {$500.00}

Test Case 2:

12/26/2017	**Mortgage**
Consumer's state GA	**Sub-product:** Conventional home mortgage
Consumer's zip 31419	**Issue** Applying for a mortgage or refinancing an existing mortgage
Submitted via Web	**Consumer consent to publish narrative** ✔ Consent provided
Did consumer dispute the response? N/A	**Consumer complaint narrative**

RE : LOAN DEPOT I received a flyer showing an attractive interest rate for a mortgage refinance so I called for further information. I answered the representative 's questions (XXXX XXXX NMLS # XXXX). And, I told him I would only be interested in a fixed rate 30-year conventional mortgage with no points or closing costs. My credit rating is more than 800 and debt to asset ratio low. He quoted me a rate very little less than what I have but more than the flyer advertised. Furthermore, there were fees exceeding {$3000.00}. I made it clear that I was NOT INTERESTED in his loan and told him to destroy my information and to stop any and all processes to apply for any loan and/or refinance. He said he would but he did not. Today I received a large stack of papers with my personal information (much wrong), and an application for a variable rate loan with even more fees that were disclosed in the phone call. I called Loan Depot today and told them to STOP THIS PROCESS, DESTROY MY INFORMATION, AND TO NOT SHARE ANY INFORMATION WITH ANYONE AT ANY TIME. The lady I spoke with said she would honor my requests. I feel betrayed, lied to, and used. My personal information has been used against my wishes and leaves me open for more abuse in the future. Loan Depot should be stopped from their deceptive practices and from abusing other people 's personal information. Thank you.

Company information

Date complaint sent to company 12/26/2017	**Timely response?** ◉ Yes
Company name LD Holdings Group, LLC	**Company response to consumer** Closed with non-monetary relief
	Company public response Company believes it acted appropriately as authorized by contract or law

App Response:

Conclusion and Next Steps

The sentiment analysis model built using BERT achieved **97% accuracy**, with strong performance across all sentiment classes. Future improvements could involve exploring model ensembles, domain-specific transformers, or leveraging more extensive datasets for further generalization.

This comprehensive pipeline—from preprocessing to deployment—provides a scalable and efficient sentiment analysis solution for customer complaints, suitable for real-world applications in customer service and feedback analysis.

Code Contribution Percentage

To assess the originality of the code used in this project, the percentage of external code copied, modified, or created from scratch is calculated.

- Total lines of code sourced from the internet: **150**

- Lines of code modified: **40**

- Lines of original code added: **90**

Percentage calculation:

Percentage of copied code = (150-40/160+90) x 100 = (110/250) x 100 ≈ 44%

Thus, approximately **44%** of the code in this project originated from external sources, with significant modifications and additions made to suit the project requirements.

References

1. Consumer Financial Protection Bureau (CFPB) Complaint Database: https://www.consumerfinance.gov/data-research/consumer-complaints/

2. Hugging Face Transformers Library: https://huggingface.co/transformers/

3. Python Streamlit Documentation: https://docs.streamlit.io/

4. NLTK VADER Sentiment Analysis: https://www.nltk.org/howto/sentiment.html